

Distributed Detection over Random Networks: Large Deviations Performance Analysis

Dragana Bajović, Dušan Jakovetić, João Xavier, Bruno Sinopoli and José M. F. Moura

Abstract

We study the large deviations performance, i.e., the exponential decay rate of the error probability, of distributed detection algorithms over random networks. At each time step k each sensor: 1) averages its decision variable with the neighbors decision variables; and 2) accounts on-the-fly for its new observation. We show that distributed detection exhibits a “phase change” behavior. When the rate of network information flow (the speed of averaging) is above a threshold, then distributed detection is asymptotically equivalent to the optimal centralized detection, i.e., the exponential decay rate of the error probability for distributed detection equals the Chernoff information. When the rate of information flow is below a threshold, distributed detection achieves only a fraction of the Chernoff information rate; we quantify this achievable rate as a function of the network rate of information flow. Simulation examples demonstrate our theoretical findings on the behavior of distributed detection over random networks.

Keywords: Chernoff information, distributed detection, random network, running consensus, information flow, large deviations.

Partially supported by grants SIPM PTDC/EEA-ACR/73749/2006 and SFRH/BD/33517/2008 (through the Carnegie Mellon/Portugal Program managed by ICTI) from Fundação para a Ciência e Tecnologia and also by ISR/IST plurianual funding (POSC program, FEDER). Work of José M. F. Moura is partially supported by NSF under grants CCF-1011903 and CCF-1018509, and by AFOSR grant FA95501010291. Dragana Bajović and Dušan Jakovetić hold fellowships from the Carnegie Mellon/Portugal Program.

Dragana Bajović and Dušan Jakovetić are with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal, and with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA dbajovic@andrew.cmu.edu, djakovet@andrew.cmu.edu

João Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal jxavier@isr.ist.utl.pt

Bruno Sinopoli and José M. F. Moura are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA brunos@ece.cmu.edu, moura@ece.cmu.edu; ph: (412) 268-6341; fax: (412) 268-3890

I. INTRODUCTION

Existing literature on distributed detection can be broadly divided into three different classes. The first studies parallel (fusion) architectures, where all sensors transmit their measurements, or local likelihood ratios, or local decisions, to a fusion node; the fusion node subsequently makes the final decision (see, e.g., [1], [2], [3], [4].) The second considers consensus-based detection, where no fusion node is required, and sensors communicate with single-hop neighbors only over a generic network (see, e.g., [5], [6]). Consensus-based detection operates in two phases. First, in the sensing phase, each sensor collects sufficient observations over a period of time. In the second, communication phase, sensors subsequently run the consensus algorithm to fuse their local log likelihood ratios. More recently, a third class of distributed detection has been proposed (see [7], [8], [9].) where, as with consensus-based detection, sensors communicate over a generic network, and no fusion node is required. Differently than consensus-based detection, sensing and communication phases occur in the same time step. Namely, at each time step k , each sensor: 1) exchanges its current decision variable with single-hop neighbors; and 2) processes its new observation, gathered at time step k .

In this paper, we focus on the third class of distributed detection, and we provide fundamental analysis of the large deviation performance, i.e., of the exponential decay rate of the error probability (as $k \rightarrow \infty$), when the underlying communication network is *randomly varying*. Namely, we show that distributed detection over random networks is asymptotically equivalent to the optimal centralized detection, if the rate of information flow (i.e., the speed of averaging) across the random network is large enough. That is, if the rate of information flow is above a threshold, then the exponential rate of decay of the error probability of distributed detection equals the Chernoff information—the best possible rate of the optimal centralized detector. When the random network has slower information flow (asymptotic optimality cannot be achieved,) we find what fraction of the best possible rate of decay of the error probability distributed detection can achieve; hence, we quantify the tradeoff between the network connectivity and achievable detection performance.

Specifically, we consider the problem where sensors cooperate over a network and sense the environment to decide between two hypothesis. The network is random, varying over time k (see, e.g., [10]); in alternative, the network uses a random communication protocol, like gossip (see, e.g., [11]). The network connectivity is described by $\{W(k)\}_{k=1}^{\infty}$, the sequence of identically distributed (i.i.d.) consensus weight matrices. The sensors' observations are Gaussian, correlated in space, and uncorrelated in time. At each time k , each sensor: 1) communicates with its single-hop neighbors to compute the weighted average of its own and the neighbors' decision variables; and 2) accounts for its new observation acquired at time

k . The network's rate of information flow (i.e., the speed of averaging,) is then measured by $|\log r|$, where r is the second largest eigenvalue of the expected value of $W(k)^2$. We then show that distributed detection exhibits a “phase change” behavior. If $|\log r|$ exceeds a threshold, then distributed detection is asymptotically optimal. If $|\log r|$ is below the threshold, then distributed detection achieves only a fraction of the best possible rate; and we evaluate the achievable rate as a function of r . Finally, we demonstrate by simulation examples our theoretical findings on the behavior of distributed detection over random networks.

Several recent references [8], [9], [7] consider different variants of distributed detection of the third class. We consider in this paper the running consensus, the variant in [7].

In the context of estimation, distributed iterative schemes have also been proposed. References [12], [13] propose diffusion type LMS and RLS algorithms for distributed estimation; references [14], [15] also propose algorithms for distributed estimation, based on the alternating direction method of multipliers. Finally, reference [16] proposes stochastic-approximation type algorithm for distributed estimation, allowing for randomly varying networks and generic (with finite second moment) observation noise. With respect to the network topology and the observation noise, we also allow for random networks, but we assume Gaussian, spatially correlated observation noise.

We comment on the differences between this work and reference [7], which also studies asymptotic performance of distributed detection via running consensus, with i.i.d. matrices $W(k)$. Reference [7] studies a problem different than ours, in which the means of the sensors' observations under the two hypothesis become closer and closer; consequently, there is an asymptotic, non zero, probability of miss, and asymptotic, non zero, probability of false alarm. Within this framework, the running consensus achieves the efficacy [17] of the optimal centralized detector, under a mild assumption on the underlying network being connected on average. In contrast, we assume that the means of the distributions do not approach each other as k grows, but stay fixed with k . The Bayes error probability exponentially decays to zero, and we examine its rate of decay. We show that, in order to achieve the optimal decay rate of the Bayes error probability, the running consensus needs an assumption *stronger* than connectedness on average, namely, the averaging speed needs to be *sufficiently large* (as measured by $|\log r|$.)

In recent work [18], we considered running consensus detection when the underlying network is *deterministically* time varying; we showed that asymptotic optimality holds if the graph that collects the union of links that are online at least once over a finite time window is connected. In contrast, we consider here the case when the underlying network or the communication protocol are *random*, and we establish a sufficient condition for optimality in terms of averaging speed (measured by $|\log r|$.)

Paper organization. The next paragraph defines notation that we use throughout the paper. Section II reviews standard asymptotic results in hypothesis testing, in particular, the Chernoff lemma. Section III explains the sensor observations model that we assume and studies the optimal centralized detection, as if there was a fusion node to process all sensors' observations. Section IV presents the running consensus distributed detection algorithm. Section V studies the asymptotic performance of distributed detection on a simple, yet illustrative, example of random matrices $W(k)$. Section VI studies asymptotic performance of distributed detection in the general case. Section VII demonstrates by simulation examples our theoretical findings. Finally, section VIII concludes the paper.

Notation. We denote by: A_{ij} or $[A]_{ij}$ (as appropriate) the (i, j) -th entry of a matrix A ; a_i or $[a]_i$ the i -th entry of a vector a ; I , 1 , and e_i , respectively, the identity matrix, the column vector with unit entries, and the i -th column of I , J the $N \times N$ matrix $J := (1/N)11^\top$; $\|\cdot\|_l$ the vector (respectively, matrix) l -norm of its vector (respectively, matrix) argument, $\|\cdot\| = \|\cdot\|_2$ the Euclidean (respectively, spectral) norm of its vector (respectively, matrix) argument, $\|\cdot\|_F$ the Frobenius norm of a matrix; $\lambda_i(\cdot)$ the i -th largest eigenvalue, $\text{Diag}(a)$ the diagonal matrix with the diagonal equal to the vector a ; $\mathbb{E}[\cdot]$ and $\mathbb{P}(\cdot)$ the expected value and probability, respectively; $\mathcal{I}_{\mathcal{A}}$ the indicator function of the event \mathcal{A} ; finally, $\mathcal{Q}(\cdot)$ the Q-function, i.e., the function that calculates the right tail probability of the standard normal distribution; $\mathcal{Q}(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-\frac{u^2}{2}} du$, $t \in \mathbb{R}$.

II. PRELIMINARIES

This section reviews standard asymptotic results in hypothesis testing, in particular, the Chernoff lemma, [19]; it also introduces certain inequalities for the $\mathcal{Q}(\cdot)$ function that we use throughout. We first formally define the binary hypothesis testing problem and the log-likelihood ratio (LLR) test.

Binary hypothesis testing problem: Log-likelihood ratio test. Consider the sequence of independent identically distributed (i.i.d.) d -dimensional random vectors (observations) $y(k)$, $k = 1, 2, \dots$, and the binary hypothesis testing problem of deciding whether the probability measure generating $y(k)$ is ν_0 (under hypothesis H_0) or ν_1 (under H_1). Assume that ν_1 and ν_0 are mutually absolutely continuous, distinguishable measures. Based on the observations $y(1), \dots, y(k)$, formally, a decision test T is a sequence of maps $T_k : \mathbb{R}^{kd} \rightarrow \{0, 1\}$, $k = 1, 2, \dots$, with the interpretation that $T_k(y(1), \dots, y(k)) = l$ means that H_l is decided, $l = 0, 1$. Specifically, consider the log-likelihood ratio (LLR) test to decide between H_0 and H_1 , where the T_k is given as follows:

$$\mathcal{D}(k) := \frac{1}{k} \sum_{j=1}^k L(k) \tag{1}$$

$$T_k = \mathcal{I}_{\{\mathcal{D}(k) > \gamma_k\}}, \tag{2}$$

where $L(k) := \log \frac{d\nu_1}{d\nu_0}(y(k))$ is the LLR (given by the Radon-Nikodym derivative of ν_1 with respect to ν_0 evaluated at $y(k)$), and γ_k is a chosen threshold.

Asymptotic Bayes detection performance: Chernoff lemma. Given a test T , we are interested in quantifying the detection performance, namely, in determining the Bayes error probability after k data (observation) samples are processed:

$$P^e(k) = P(H_0) \alpha(k) + P(H_1) \beta(k), \quad (3)$$

where $P(H_i)$ are the prior probabilities, $\alpha(k) = \mathbb{P}(T_k = 1|H_0)$ and $\beta(k) = \mathbb{P}(T_k = 0|H_1)$ are, respectively, the probability of false alarm and the probability of a miss. Generally, exact evaluation of $\alpha(k)$ and $\beta(k)$ (and hence, $P^e(k)$) is very hard (as in the case of distributed detection over random networks that we study; see also [20] for distributed detection on a parallel architecture.) We seek computationally tractable estimates of $P^e(k)$, when k grows large. Typically, for large k , $P^e(k)$ is a small number (i.e., the detection error occurs rarely,) and, in many models, it exponentially decays to zero as $k \rightarrow +\infty$. Thus, it is of interest to determine the (large deviations) rate of exponential decay of $P^e(k)$, given by:

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P^e(k). \quad (4)$$

Lemma 1 ([19], [21]) states that, among all possible decision tests, the LLR test with zero threshold maximizes (4) (i.e., has the fastest decay rate of $P^e(k)$.) The corresponding decay rate equals the Chernoff information \mathbf{C} , i.e., the Chernoff distance between the distributions of $y(k)$ under H_0 and H_1 , where \mathbf{C} is given by, [19]:

$$\mathbf{C} = \max_{s \in [0,1]} \left\{ -\log \int \left(\frac{d\nu_0}{d\nu_1} \right)^{1-s} d\nu_1 \right\}. \quad (5)$$

Lemma 1 (Chernoff lemma) If $P(H_0) \in (0, 1)$, then:

$$\sup_T \left\{ \limsup_{k \rightarrow \infty} -\frac{1}{k} \log P^e(k) \right\} = \mathbf{C}, \quad (6)$$

where the supremum over all possible tests T is attained for the LLR test with $\gamma_k = 0$, $\forall k$.

Asymptotically optimal test. We introduce the following definition of the asymptotically optimal test.

Definition 2 The decision test T is asymptotically optimal if it attains the supremum in eqn. (6).

We will find a necessary condition and a sufficient condition for asymptotic optimality (in the sense of Definition 2) of the running consensus distributed detection.

Inequalities for the standard normal distribution. We will use the following property of the $\mathcal{Q}(\cdot)$ function, namely, that for any $t > 0$ (e.g., [22]):

$$\frac{t}{1+t^2} e^{-\frac{t^2}{2}} \leq \sqrt{2\pi} \mathcal{Q}(t) \leq \frac{1}{t} e^{-\frac{t^2}{2}}. \quad (7)$$

III. CENTRALIZED DETECTION

We proceed with the Gaussian model for which we find (in section V) conditions for asymptotic optimality of the running consensus distributed detection. Subsection III-A describes the model of the sensor observations that we assume. Subsection III-B describes the (asymptotically) optimal centralized detection, as if there was a fusion node that collects and processes the observations from all sensors.

A. Sensor observations model

We assume that N sensors are deployed to sense the environment and to decide between the two possible hypothesis, H_1 and H_0 . Each sensor i measures a scalar quantity $y_i(k)$ at each time step k ; all sensors measure at time steps $k = 1, 2, \dots$. Collect $y_i(k)$'s, $i = 1, \dots, N$, into $N \times 1$ vector $y(k) = (y_1(k), \dots, y_N(k))^\top$. We assume that $\{y(k)\}$ has the following distribution:

$$\text{Under } H_l : y(k) = m_l + \zeta(k), \quad l = 0, 1. \quad (8)$$

The quantity m_l is the constant signal; the quantity $\zeta(k)$ is zero-mean, Gaussian, spatially correlated noise, i.i.d. across time, with distribution $\zeta(k) \sim \mathcal{N}(0, S)$, where S is a positive definite covariance matrix. Spatial correlation of the measurements (i.e., non-diagonal covariance matrix S) accounts for, e.g., dense deployment in sensor networks.

B. (Asymptotically) optimal centralized detection

This subsection studies optimal centralized detection under the Gaussian assumptions in III-A, as if there was a fusion node that collects and processes all sensors' observations. The LLR decision test is given by eqns. (1) and (2), where it is straightforward to show that now the LLR takes the following form:

$$L(k) = (m_1 - m_0)^\top S^{-1} \left(y(k) - \frac{m_1 + m_0}{2} \right) \quad (9)$$

Conditioned on either hypothesis H_1 and H_0 , $L(k) \sim \mathcal{N}\left(m_L^{(l)}, \sigma_L^2\right)$, where

$$m_L^{(1)} = -m_L^{(0)} = \frac{1}{2}(m_1 - m_0)^\top S^{-1}(m_1 - m_0) \quad (10)$$

$$\sigma_L^2 = (m_1 - m_0)^\top S^{-1}(m_1 - m_0). \quad (11)$$

Define the vector $v \in \mathbb{R}^N$ as

$$v := S^{-1}(m_1 - m_0). \quad (12)$$

Then, the LLR $L(k)$ can be written as follows:

$$L(k) = \sum_{i=1}^N v_i \left(y_i(k) - \frac{[m_1]_i + [m_0]_i}{2} \right) = \sum_{i=1}^N \eta_i(k) \quad (13)$$

Thus, the LLR at time k is separable, i.e., the LLR is the sum of the terms $\eta_i(k)$ that depend affinely on the individual observations $y_i(k)$. We will exploit this fact in section IV to derive the distributed, running consensus, detection algorithm.

Bayes probability of error: finite number of observations. The minimal Bayes error probability, $P_{\text{cen}}^e(k)$, when k samples are processed, and $P(H_0) = P(H_1) = \frac{1}{2}$ (equal prior probabilities), is attained for the (centralized) LLR test with zero threshold; $P_{\text{cen}}^e(k)$ equals:

$$P_{\text{cen}}^e(k) = \mathcal{Q} \left(\sqrt{k} \frac{m_L^{(1)}}{\sigma_L} \right). \quad (14)$$

The quantity $P_{\text{cen}}^e(k)$ will be of interest when we compare (by simulation, in Section VII) the running consensus detection with the optimal centralized detection, in the regime of finite k .

Bayes probability of error: time asymptotic results. The Chernoff lemma (Lemma 1) applies also to the (centralized) detection problem as defined in subsection III-A. It can be shown that the Chernoff information, in this case, equals:

$$\mathbf{C} = \mathbf{C}_{\text{tot}} = \frac{1}{8} (m_1 - m_0)^\top S^{-1} (m_1 - m_0). \quad (15)$$

In eqn. (15), the subscript tot designates the total Chernoff information of the network, i.e., the Chernoff information of the observations collected from all sensors. Specifically, if the sensor observations are uncorrelated (the noise covariance matrix $S = \text{Diag}(\sigma_1^2, \dots, \sigma_N^2)$), then:

$$\mathbf{C}_{\text{tot}} = \sum_{i=1}^N \frac{[m_1 - m_0]_i^2}{8\sigma_i^2} = \sum_{i=1}^N \mathbf{C}_i, \quad (16)$$

where \mathbf{C}_i is the Chernoff information of the individual sensor i . That is, \mathbf{C}_i equals the best achievable rate of the Bayes error probability, if the sensor i worked as an individual (it did not cooperate with the other sensors.)

Lemma 3 (Chernoff lemma for asymptotically optimal centralized detector) Consider the observation model defined in subsection III-A, and let $P(H_0) \in (0, 1)$. The LLR test with $\gamma_k = 0, \forall k$, is asymptoti-

cally optimal in the sense of Definition 2. Moreover, for the LLR test with $\gamma_k = 0, \forall k$, we have:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log P_{\text{cen}}^e(k) = -\mathbf{C}_{\text{tot}}, \quad (17)$$

where \mathbf{C}_{tot} is given by eqn. (15).

IV. DISTRIBUTED DETECTION

We now consider distributed detection, under the same assumptions on the sensor observations as in III-A; but the fusion node is no longer available, and the sensors cooperate through a randomly varying network. Specifically, we consider the running consensus distributed detection, proposed in [7], and we extend it to spatially correlated observations. At each time k , each sensor i improves its decision variable, call it $x_i(k)$, two-fold: 1) by exchanging the decision variable locally with its neighbors and computing the weighted average of its own and the neighbors' variables; and 2) by incorporating its new observation at time k .

Recall the definition of the vector v in eqn. (12) and the scalar $\eta_i(k)$ in eqn. (13). The update of $x_i(k)$ is then as follows:

$$\begin{aligned} x_i(k+1) &= \frac{k}{k+1} \left(W_{ii}(k)x_i(k) + \sum_{j \in O_i(k)} W_{ij}(k)x_j(k) \right) + \frac{1}{k+1} \eta_i(k+1), \quad k = 1, \dots \\ x_i(1) &= \eta_i(1). \end{aligned} \quad (18)$$

Here $O_i(k)$ is the (random) neighborhood of sensor i at time k , and $W_{ij}(k)$ are the (random) averaging weights.¹ The local sensor i 's decision test at time k , $T_{k,i}$, is given by:

$$T_{k,i} := \mathcal{I}_{\{x_i(k) > 0\}}, \quad (19)$$

i.e., H_1 (resp. H_0) is decided when $x_i(k) > 0$ (resp. $x_i(k) \leq 0$.) Let $x(k) = (x_1(k), x_2(k), \dots, x_N(k))^T$ and $\eta(k) = (\eta_1(k), \dots, \eta_N(k))^T$. Also, collect the averaging weights $W_{ij}(k)$ in $N \times N$ matrix $W(k)$, where, clearly, $W_{ij}(k) = 0$ if the sensors i and j do not communicate at time step k . The algorithm in matrix form becomes:

$$\begin{aligned} x(k+1) &= \frac{k}{k+1} W(k)x(k) + \frac{1}{k+1} \eta(k+1), \quad k = 1, \dots \\ x(1) &= \eta(1). \end{aligned} \quad (20)$$

¹We remark that, to implement the algorithm, sensor i has to know the quantities $v_i := [S^{-1}(m_1 - m_0)]_i$, $[m_1]_i$ and $[m_0]_i$; this knowledge can be acquired in the training period of the sensor network.

We remark that the algorithm in (20) extends the running consensus algorithm in [23] for spatially correlated sensor observations (non-diagonal covariance matrix S .) When S is diagonal, the algorithm in (20) reduces to the algorithm in [23].²

We allow the averaging matrices $W(k)$ to be random. Formally, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space (where Ω is a sample space, \mathcal{F} is a σ -algebra, and $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ is a probability measure.) For any k , $W(k)$ is a random variable, i.e., an \mathcal{F} -measurable function $W(k) = W(\omega; k)$, $\omega \in \Omega$, $W(k) : \Omega \rightarrow \mathbb{R}^{N \times N}$. We now summarize the assumptions on $W(k)$. Recall that $J := \frac{1}{N}(11^\top)$ and denote by $\widetilde{W}(k) := W(k) - J$. From now on, we will drop the index k from $W(k)$ and $\widetilde{W}(k)$ when we refer to the distribution of $W(k)$ and $\widetilde{W}(k)$.

Assumption 4 For the sequence of matrices $\{W(k)\}_{k=1}^\infty$, we assume the following:

- 1) The sequence $\{W(k)\}_{k=1}^\infty$ is i.i.d.
- 2) W is symmetric and stochastic (row-sums are equal to 1 and the entries are nonnegative,) with probability one.
- 3) The random matrix $W(l)$ and the random vector $y(s)$ are independent, $\forall l, \forall s$.

In sections V and VI, we examine what (additional) conditions the matrices $W(k)$ have to satisfy, to achieve asymptotic optimality of the distributed detection algorithm.

Network supergraph. Define also the network supergraph as a pair $G := (\mathcal{V}, E)$, where \mathcal{V} is the set of nodes with cardinality $|\mathcal{V}| = N$, and E is the set of edges with cardinality $|E| = M$, defined by: $E = \{\{i, j\} : \mathbb{P}(W_{ij} \neq 0) > 0, i < j\}$. Clearly, when, for some $\{i, j\}$, $\mathbb{P}(W_{ij} \neq 0) = 0$, then the link $\{i, j\} \notin E$ and nodes i and j never communicate.

For subsequent analysis, it will be useful to define the matrices $\Phi(k, j)$, for $k > j \geq 1$, as follows:

$$\Phi(k, j) := W(k-1)W(k-2)\dots W(j). \quad (21)$$

Then, the algorithm in eqn. (20) can be written as:

$$x(k) = \frac{1}{k} \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) + \frac{1}{k} \eta(k), \quad k = 2, 3, \dots \quad (22)$$

Also, introduce:

$$\widetilde{\Phi}(k, j) := \widetilde{W}(k-1)\widetilde{W}(k-2)\dots\widetilde{W}(j), \quad k > j \geq 1, \quad (23)$$

²Another minor difference between [23] and eqn. (20) is that [23] multiplies the log-likelihood ratio term (the term analogous to $\frac{1}{k+1}\eta(k+1)$) by N ; this multiplication does not affect detection performance.

and remark that

$$\tilde{\Phi}(k, j) = \Phi(k, j) - J.$$

Recall the definition of the $N \times 1$ vector v in (12). The sequence of $N \times 1$ random vectors $\{\eta(k)\}$, conditioned on H_l , is i.i.d. The vector $\eta(k)$ (under hypothesis H_l , $l = 0, 1$) is Gaussian with mean $m_\eta^{(l)}$ and covariance S^η :

$$m_\eta^{(l)} = (-1)^{(l+1)} \text{Diag}(v) \frac{1}{2}(m_1 - m_0) \quad (24)$$

$$S^\eta = \text{Diag}(v) S \text{Diag}(v). \quad (25)$$

Here $\text{Diag}(v)$ is a diagonal matrix with the diagonal entries equal to the entries of v .

V. ASYMPTOTIC PERFORMANCE OF DISTRIBUTED DETECTION: SWITCHING FUSION EXAMPLE

In this section, we examine asymptotic performance of distributed detection algorithm on a simple and impractical, yet illustrative example; we tackle the generic case in Section VI. The network at a time step k can either be fully connected, with probability p , or completely disconnected (without edges,) with probability $1 - p$. Specifically, the distribution of the random averaging matrix $W(k)$ is given by:

$$W(k) = \begin{cases} J & \text{with prob. } p \\ I & \text{with prob. } 1 - p. \end{cases} \quad (26)$$

With model (26), at each time step k , each sensor behaves as a fusion node, with probability p , and as an individual detector, with probability $1 - p$. We call this communication model the switching fusion. We show that, to achieve asymptotic optimality of distributed detection, the fusion step ($W(k) = J$) should occur sufficiently often, i.e., p should exceed a threshold. Namely, we find necessary and sufficient condition for the asymptotic optimality in terms of p . When distributed detection is not optimal (p is below the threshold,) we find the achievable rate of decay of the error probability, as a function of p . The goal of the switching fusion example is two-fold. First, it provides insight on how the amount of communication (measured by p) affects detection performance. Second, it explains in a clear and natural way our methodology for quantifying detection performance on generic networks (in Section VI.) Namely, Section VI mimics and extends the analysis from Section V to derive distributed detection performance on generic networks. We next detail the sensor observations model.

We assume that the observations $y_i(k)$ of N different sensors are uncorrelated, and that the individual Chernoff information, given by eqn. (16), is the same at each sensor i . Hence, we have $\mathbf{C}_{\text{tot}} = N \mathbf{C}_i$. We assume that, at time instant k , the network can either be fully connected, with probability p , or without

edges, with probability $1 - p$.

Denote by $P_{i,\text{dis}}^e(k)$ the Bayes error probability at sensor i , after k samples are processed. We have the following Theorem on the asymptotic performance of the distributed detection algorithm.

Theorem 5 Consider the distributed detection algorithm given by eqns. (18) and (19). Assume that the sensor observations are spatially uncorrelated and that the Chernoff information \mathbf{C}_i is equal at each sensor i . Let $W(k)$ be i.i.d. matrices with the distribution given by eqn. (26). Then, the exponential decay rate of the error probability is given by:

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{i,\text{dis}}^e(k) = \begin{cases} \mathbf{C}_{\text{tot}} & \text{if } |\log(1-p)| \geq \mathbf{C}_{\text{tot}}(N-1) \\ \mathbf{C}_i + |\log(1-p)| & \text{if } |\log(1-p)| \leq \frac{\mathbf{C}_{\text{tot}}(N-1)}{N^2} \\ 2\sqrt{\frac{|\log(1-p)|\mathbf{C}_{\text{tot}}}{N-1}} - \frac{|\log(1-p)|}{N-1} & \text{otherwise.} \end{cases} \quad (27)$$

Moreover, a necessary and sufficient condition for asymptotic optimality, in the sense of Definition 2, is given by:

$$\frac{|\log(1-p)|}{N-1} \geq \mathbf{C}_{\text{tot}} = N\mathbf{C}_i. \quad (28)$$

Condition (28) says that the network connectivity should be good enough (i.e., p should be large enough,) in order to achieve the asymptotic optimality of distributed detection. Also, there is a “phase change” behavior, in a sense that distributed detection is asymptotically optimal above a threshold on p , and it is not optimal below that threshold. Further, we can see that, as p decreases, distributed detection performance becomes worse and worse, and it approaches the performance of an individual sensor-detector. (See Figure 1, and eqn. (27).)

We proceed with proving Theorem 5. In Section VI, we will follow a reasoning similar to the proof of Theorem 5 to provide a sufficient condition for asymptotic optimality on generic networks.

Proof of Theorem 5: First, remark that $x(k)$, conditioned on H_0 , is equal in distribution to $-x(k)$, conditioned on H_1 . This is true because $\eta(k)$, conditioned on H_0 , is equal in distribution to $-\eta(k)$, conditioned on H_1 , for all k ; and the distribution of $W(k)$ does not depend on the active hypothesis, H_0 or H_1 . Denote by $\mathbb{P}_l(\cdot) = \mathbb{P}(\cdot|H_l)$, $l = 0, 1$, and consider the probability of false alarm, the probability of miss, and the Bayes error probability at sensor i (with the running consensus detector,) respectively, given by:

$$\alpha_{i,\text{dis}}(k) = \mathbb{P}_0(x_i(k) > 0), \quad \beta_{i,\text{dis}}(k) = \mathbb{P}_1(x_i(k) \leq 0) \quad (29)$$

$$P_{i,\text{dis}}^e(k) = P(H_0)\mathbb{P}_0(x_i(k) > 0) + P(H_1)\mathbb{P}_1(x_i(k) \leq 0). \quad (30)$$

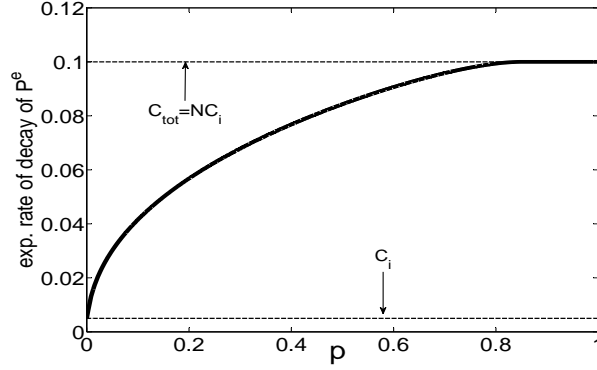


Fig. 1. Exponential decay rate of error probability ϕ^* (given by eqn. (26)) for the model considered in Section V. The network has $N = 20$ sensors and $C_{\text{tot}} = 0.1$. The optimal rate (equal to C_{tot}) is achieved for $p \geq 0.83$.

Remark that $\mathbb{P}_l(x_i(k) = 0) = 0$, $l = 0, 1$. Thus, we have that

$$\beta_{i,\text{dis}}(k) = \mathbb{P}_1(x_i(k) \leq 0) = \mathbb{P}_0(x_i(k) > 0) = \alpha_{i,\text{dis}}(k), \quad \forall k, \forall i. \quad (31)$$

From eqns. (30) and (31), it can be shown that:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) = \limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) \quad (32)$$

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) = \liminf_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k). \quad (33)$$

We further assume that H_0 is true, and we restrict our attention to $\alpha_{i,\text{dis}}(k)$, but the same conclusions (from (31)) will be valid for $\beta_{i,\text{dis}}(k)$ also. We now make the key step in proving Theorem 5, by defining a partition of the probability space Ω . Fix the time step k and denote by A_l , $l = 0, \dots, k-1$, the event

$$A_l = \begin{cases} \{\max\{s \in \{1, \dots, k-1\} : W(s) = J\} = l\} & \text{for } l = 1, \dots, k-1 \\ \{W(s) = I, s = 1, \dots, k-1\} & \text{for } l = 0. \end{cases}$$

That is, A_l is the event that the largest time step $s \leq k-1$, for which $W(s) = J$, is equal to $s = l$. (The event A_l includes the scenarios of arbitrary realizations of $W(s)$ —either J or I —for $s \leq l$; but it requires $W(s) = I$ for all $l < s \leq k-1$.) Remark that A_l is a function of k , but the dependence on k is dropped for notation simplicity. We have that $\mathbb{P}(A_l) = p(1-p)^{k-l-1}$, for $l = 1, \dots, k-1$, and $\mathbb{P}(A_0) = (1-p)^{k-1}$. Also, each two events, A_l and A_j , $j \neq l$, are disjoint, and $\cup_{l=0}^{k-1} A_l = \Omega$, i.e., the events A_l , $l = 0, \dots, k-1$, constitute a finite partition of the probability space Ω . (Note that

$\sum_{l=0}^{k-1} \mathbb{P}(A_l) = 1$.) Recall the definition of $\Phi(k, j)$ in eqn. (21) and note that, if A_l occurred, we have:

$$\Phi(k, s) = \begin{cases} I & \text{if } k-1 \geq s > l \\ J & \text{if } s \leq l. \end{cases} \quad (34)$$

Further, conditioned on A_l , we have that (when $l = 0$, first sum in eqn. (35) does not exist)

$$\begin{aligned} x(k) &= \frac{1}{k} \left(\sum_{j=1}^l J\eta(j) + \sum_{j=l+1}^k I\eta(j) \right) \\ &= \frac{1}{k} \left(\sum_{j=1}^l \left(\frac{1^\top \eta(j)}{N} \right) 1 + \sum_{j=l+1}^k \eta(j) \right). \end{aligned} \quad (35)$$

Hence, conditioned on A_l , $x_i(k)$ is a Gaussian random variable,

$$x_i(k)|A_l \sim \mathcal{N}(\theta(l; k), \zeta^2(l; k)),$$

where

$$\theta(l; k) = -\frac{4l}{N} \mathbf{C}_{\text{tot}} - 4(k-l) \mathbf{C}_i = -4k \mathbf{C}_i \quad (36)$$

$$\zeta^2(l; k) = \frac{8l}{N^2} \mathbf{C}_{\text{tot}} + 8(k-l) \mathbf{C}_i. \quad (37)$$

Define

$$\chi(l; k) := -\frac{\theta(l; k)}{\zeta(l; k)} = \frac{\sqrt{2\mathbf{C}_i} k}{\sqrt{\frac{l}{N} + (k-l)}}, \quad (38)$$

and remark that

$$\mathbb{P}_0(x_i(k) > 0 | A_l) = \mathcal{Q}(\chi(l; k)),$$

where $\mathbb{P}_0(\cdot) := \mathbb{P}(\cdot | H_0)$.

Using the total probability, we can write $\alpha_{i,\text{dis}}(k) = \mathbb{P}_0(x_i(k) > 0)$ as:

$$\begin{aligned} \alpha_{i,\text{dis}}(k) &= \sum_{l=0}^{k-1} \mathbb{P}_0(x_i(k) > 0 | A_l) \mathbb{P}(A_l) \\ &= \sum_{l=1}^{k-1} \mathcal{Q}(\chi(l; k)) p(1-p)^{k-l-1} + \mathcal{Q}(\chi(0; k)) (1-p)^{k-1}. \end{aligned} \quad (39)$$

We now proceed with calculating the exponential rate of decay of $\alpha_{i,\text{dis}}(k)$ (and hence, $P_{i,\text{dis}}^e(k)$) as $k \rightarrow \infty$. The key ingredient to do that is the representation of $\alpha_{i,\text{dis}}(k)$ in eqn. (39), and the inequalities for the \mathcal{Q} -function in eqn. (7). Namely, it can be shown that, when k grows large, $\alpha_{i,\text{dis}}(k)$, and hence,

$P_{i,\text{dis}}^e(k)$, behaves as (here we present the main idea but precise statements are in the Appendix):

$$P_{i,\text{dis}}^e(k) \sim \sum_{j=0}^{k-1} e^{-k \phi(j;k)}, \quad (40)$$

where

$$\phi(j;k) := \frac{\mathbf{C}_{\text{tot}}}{1 + (N-1)\frac{j+1}{k}} + \frac{j}{k} |\log(1-p)|. \quad (41)$$

Hence, the quantities $\phi(j;k)$, for different j 's, represent different “modes” of decay; the decay of $P_{i,\text{dis}}^e(k)$ is then determined by the slowest mode $\widehat{\phi}(k)$, defined by:

$$\widehat{\phi}(k) = \min_{j=0,\dots,k-1} \phi(j;k). \quad (42)$$

More precisely, using eqns. (32), (33), the expression for $\alpha_{i,\text{dis}}(k)$ in eqn. (39), and the inequalities (7), it can be shown that:

$$\begin{aligned} \liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_{i,\text{dis}}^e(k) &\geq \liminf_{k \rightarrow \infty} \widehat{\phi}(k) \\ \limsup_{k \rightarrow \infty} -\frac{1}{k} \log P_{i,\text{dis}}^e(k) &\leq \limsup_{k \rightarrow \infty} \widehat{\phi}(k). \end{aligned} \quad (43)$$

The detailed proof of inequalities (43) is in the Appendix.

We proceed by noting that the minimum of $\phi(j;k)$ over the discrete set $j \in \{0, 1, \dots, k-1\}$ does not differ much from the minimum of $\phi(j;k)$ over the interval $[0, k-1]$. Denote by $\phi^*(k)$ the minimum of $\phi(j;k)$ over $[0, k-1]$:

$$\phi^*(k) = \min_{j \in [0, k-1]} \phi(j;k). \quad (44)$$

Then, it is easy to verify that:

$$\phi^*(k) \leq \widehat{\phi}(k) \leq \phi^*(k) \left(1 + \frac{N-1}{k}\right) + \frac{|\log(1-p)|}{k}. \quad (45)$$

The function $\phi(j,k)$ is convex in its first argument on $j \in [0, k-1]$; it is straightforward to calculate $\phi^*(k)$, which can be shown to be equal to:

$$\phi^*(k) = \begin{cases} \frac{\mathbf{C}_{\text{tot}}}{1 + \frac{N-1}{k}} & \text{if } |\log(1-p)| \geq \frac{\mathbf{C}_{\text{tot}}(N-1)}{(1 + \frac{N-1}{k})^2} \\ \frac{\mathbf{C}_{\text{tot}}}{N} + \frac{k-1}{k} |\log(1-p)| & \text{if } |\log(1-p)| \leq \frac{\mathbf{C}_{\text{tot}}(N-1)}{N^2} \\ 2\sqrt{\frac{|\log(1-p)|\mathbf{C}_{\text{tot}}}{N-1}} - \frac{|\log(1-p)|}{N-1} - \frac{|\log(1-p)|}{k} & \text{otherwise.} \end{cases} \quad (46)$$

The limit $\lim_{k \rightarrow \infty} \phi^*(k) =: \phi^*$ exists, and is equal to:

$$\phi^* = \begin{cases} \mathbf{C}_{\text{tot}} & \text{if } |\log(1-p)| \geq \mathbf{C}_{\text{tot}}(N-1) \\ \frac{\mathbf{C}_{\text{tot}}}{N} + |\log(1-p)| & \text{if } |\log(1-p)| \leq \frac{\mathbf{C}_{\text{tot}}(N-1)}{N^2} \\ 2\sqrt{\frac{|\log(1-p)|\mathbf{C}_{\text{tot}}}{N-1}} - \frac{|\log(1-p)|}{N-1} & \text{otherwise.} \end{cases} \quad (47)$$

From eqns. (45) and (47), we have:

$$\lim_{k \rightarrow \infty} \widehat{\phi}(k) = \lim_{k \rightarrow \infty} \phi^*(k) = \phi^*. \quad (48)$$

In view of eqns. (48) and (43), it follows that the rate of decay of the error probability at sensor i is:

$$\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{i,\text{dis}}^e(k) = \lim_{k \rightarrow \infty} \widehat{\phi}(k) = \phi^*. \quad (49)$$

The necessary and sufficient condition for asymptotic optimality then follows from eqn. (47). ■

VI. ASYMPTOTIC PERFORMANCE OF DISTRIBUTED DETECTION: GENERAL CASE

This section provides a necessary condition, and a sufficient condition for asymptotic optimality of distributed detection on generic networks and for generic, spatially correlated, Gaussian observations. When distributed detection is not guaranteed to be optimal, this section finds a lower bound on the exponential decay rate of error probability, in terms of the system parameters. We start by pursuing sufficient conditions for optimality and evaluating the lower bound on the decay rate of the error probability.

A. Sufficient condition for asymptotic optimality

Recall that $r := \lambda_2(\mathbb{E}[W(k)^2]) = \|\mathbb{E}[W(k)^2] - J\|$. It is well known that the quantity r measures the speed of the information flow, i.e., the speed of the averaging across the network, like with standard consensus and gossip algorithms, e.g., [11]. (The smaller r is, the faster the averaging is.) The next Theorem shows that distributed detection is asymptotically optimal if the network information flow is fast enough, i.e., if r is small enough. The Theorem also finds a lower bound on the rate of decay of the error probability, even when the sufficient condition for asymptotic optimality does not hold. Recall also \mathbf{C}_{tot} in eqn. (16).

Theorem 6 Let Assumption 4 hold and consider the distributed detection algorithm defined by eqns. (18) and (19). Then, the following holds for the exponential decay rate of the error probability at each sensor:

$$\liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_{i,\text{dis}}^e(k) \geq \begin{cases} \mathbf{C}_{\text{tot}} & \text{if } |\log r| \geq \frac{1}{8} N^2 \left(1 + \left(1 - \frac{1}{N}\right) K\right) \|S^\eta\| \\ -\left(\frac{1}{2N^2} \sigma_L^2 \bar{\mu}^2 + \frac{1}{N} m_L^{(0)} \bar{\mu}\right) & \text{otherwise,} \end{cases}$$

where

$$\bar{\mu} = \begin{cases} \frac{1}{4} \frac{K}{K+1} + \frac{1}{4} \frac{\sqrt{K^2 + \frac{32|\log r|}{\|S^\eta\|}} (1+K)}{K+1}, \\ \text{if } \frac{1}{8} \|S^\eta\| < |\log r| < \frac{1}{8} N^2 \left(1 + \left(1 - \frac{1}{N}\right) K\right) \|S^\eta\|; \\ \frac{1}{4} \sqrt{K^2 + \frac{32|\log r|}{\|S^\eta\|}} - \frac{1}{4} K, \\ \text{if } |\log r| \leq \frac{1}{8} \|S^\eta\|. \end{cases} \quad (50)$$

Here $K = (8\bar{m}) / \|S^\eta\|$. Moreover, each sensor i is asymptotically optimal, and $\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{i,\text{dis}}^e(k) = \mathbf{C}_{\text{tot}}, \forall i$, provided that:

$$|\log r| \geq \frac{1}{8} N^2 \left(1 + \left(1 - \frac{1}{N}\right) K\right) \|S^\eta\|. \quad (51)$$

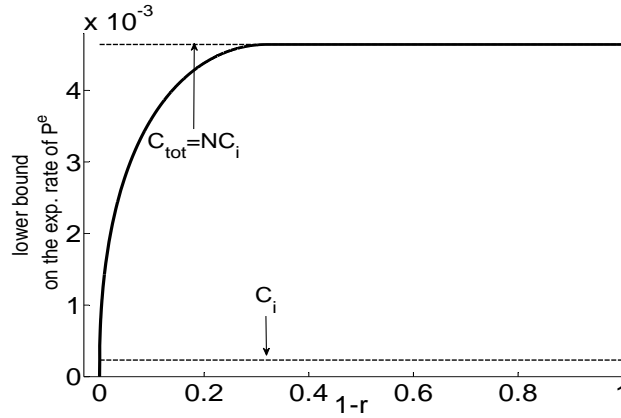


Fig. 2. Lower bound on the exponential decay rate of the error probability (given by eqn. (52)) for the (generic) model in Section VI. The network has $N = 20$ sensors, and $\mathbf{C}_{\text{tot}} = 0.0047$. The optimal rate (equal to \mathbf{C}_{tot}) is achieved for $p \geq 0.29$.

We proceed by proving Theorem 6. We first set up the proof and state some auxiliary Lemmas. We first consider the probability of false alarm, $\alpha_{i,\text{dis}}(k) = \mathbb{P}_0(x_i(k) > 0)$, but the same conclusions will hold for $\beta_{i,\text{dis}}(k) = \mathbb{P}_1(x_i(k) < 0)$ also (See the Proof of Theorem 5.) We examine the family of Chernoff bounds on the probability of false alarm, parametrized by $\mu > 0$, given by:

$$\alpha_{i,\text{dis}}(k) \leq \mathbb{E}_0 \left[e^{k\mu x_i(k)} \right] =: \mathcal{C}(k\mu), \quad (52)$$

where $\mathbb{E}_l[a] := \mathbb{E}[a|H_l]$, $l = 0, 1$. We then examine the conditions under which the best Chernoff bound falls below the negative of the Chernoff information, in the limit as $k \rightarrow \infty$. More precisely, we examine under which conditions the following inequality holds:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{C}(k\mu) \leq -\mathbf{C}_{\text{tot}}. \quad (53)$$

In subsequent analysis, we will use the following Lemmas, Lemma 7 and Lemma 8; proof of Lemma 7 is trivial, while proof of Lemma 8 is in the Appendix.

Lemma 7 Let V be a $N \times N$ stochastic matrix and consider the matrix $\tilde{V} = V - J$. Then, for all $i = 1, \dots, N$, the following inequalities hold:

$$\sum_{l=1}^N |\tilde{V}_{il}| \leq 2 \frac{N-1}{N} < 2 \quad (54)$$

$$\sum_{l=1}^N |\tilde{V}_{il}|^2 \leq \frac{N-1}{N} < 1. \quad (55)$$

Lemma 8 Let Assumption 4 hold. Then, the following inequality holds:

$$\mathbb{P} \left(\|\tilde{\Phi}(k, j)\| > \epsilon \right) \leq \frac{N^4}{\epsilon^2} r^{k-j}. \quad (56)$$

When proving Theorem 6, we will partition the time interval from the first to $(k-1)$ -th time step in windows of width B , for some integer $B \geq 1$. That is, we consider the subsets of consecutive time steps $\{k-B, k-B+1, \dots, k-1\}$, $\{k-2B, \dots, k-B-1\}$, ..., $\{k-(\mathcal{J}-1)B, \dots, k-\mathcal{J}B-1\}$, $\{1, 2, \dots, k-\mathcal{J}B-1\}$, where \mathcal{J} is the integer part of $\frac{k-1}{B}$. (Note that the total number of these subsets is $\mathcal{J}+1$; each of these subsets contains B time steps, except $\{1, 2, \dots, k-\mathcal{J}B-1\}$ that in general has the number of time steps less or equal to B .) We then define the events \mathcal{A}_j , $j = 1, \dots, \mathcal{J}+1$, as follows:

$$\begin{aligned} \mathcal{A}_1 &= \left\{ \|\tilde{\Phi}(k, k-B)\| \leq \epsilon \right\} \\ \mathcal{A}_2 &= \left\{ \|\tilde{\Phi}(k, k-B)\| > \epsilon \right\} \cap \left\{ \|\tilde{\Phi}(k, k-2B)\| \leq \epsilon \right\} \\ &\vdots \\ \mathcal{A}_{j+1} &= \left\{ \|\tilde{\Phi}(k, k-jB)\| > \epsilon \right\} \cap \left\{ \|\tilde{\Phi}(k, k-(j+1)B)\| \leq \epsilon \right\} \\ &\vdots \\ \mathcal{A}_{\mathcal{J}} &= \left\{ \|\tilde{\Phi}(k, k-(\mathcal{J}-1)B)\| > \epsilon \right\} \cap \left\{ \|\tilde{\Phi}(k, k-\mathcal{J}B)\| \leq \epsilon \right\} \\ \mathcal{A}_{\mathcal{J}+1} &= \left\{ \|\tilde{\Phi}(k, k-\mathcal{J}B)\| > \epsilon \right\}. \end{aligned} \quad (57)$$

It is easy to see that the events \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$, constitute a finite partition of the probability space Ω (i.e., each \mathcal{A}_i and \mathcal{A}_j , $i \neq j$, are disjoint, and the union of \mathcal{A}_j 's, $j = 1, \dots, \mathcal{J} + 1$, is Ω .)

By Lemma 8, the probability of the event \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$, is bounded from above as follows:

$$\mathbb{P}(\mathcal{A}_j) \leq \frac{N^4}{\epsilon^2} r^{(j-1)B}. \quad (58)$$

We will next explain the idea behind using the partition \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$. First, we state Lemma 9 that helps to understand this idea.

Lemma 9 Consider the random matrices $\tilde{\Phi}(k, j)$ given by eqn. (23), and fix two time steps, k and j , $k > j$. Suppose that the event \mathcal{A}_j occurred. Then, for any $s \leq k - jB$, and for $i = 1, \dots, N$:

$$\sum_{l=1}^N |[\tilde{\Phi}(k, s)]_{il}| \leq N\sqrt{N}\epsilon \quad (59)$$

$$\sum_{l=1}^N |[\tilde{\Phi}(k, s)]_{il}|^2 \leq N\epsilon^2. \quad (60)$$

Proof: See the Appendix. ■

Consider now the partition \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$. Given that \mathcal{A}_j occurred, the matrices $\Phi(k, s)$, for $s \leq k - jB$, are “ ϵ -close” to J (in view of Lemma 9.) For $s > k - jB$, $\Phi(k, s)$ is not “ ϵ -close” to J , and thus we pass to the “worst case” as given by Lemma 7. (We may think of this as setting $\Phi(k, s) \approx I$.) Hence, given that \mathcal{A}_j occurred, $x(k)$, in a crude approximation, behaves as:

$$x(k) \sim \frac{1}{k} \sum_{l=1}^{k-jB} J\eta(l) + \frac{1}{k} \sum_{l=k-jB+1}^k I\eta(l).$$

Then, by conditioning $x(k)$ on \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$, and by using the total probability (as with Theorem 5,) we express $P_{i,\text{dis}}^e(k)$ in terms of the $j = 1, \dots, \mathcal{J} + 1$ “modes of decay” that correspond to \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$. Finally, we find the slowest among the $\mathcal{J} + 1$ modes, as with Theorem 5. The difference compared to Theorem 5, however, is that here we work with the Chernoff bounds on $\alpha_{i,\text{dis}}(k)$ and $P_{i,\text{dis}}^e(k)$, rather than directly with $\alpha_{i,\text{dis}}(k)$ and $P_{i,\text{dis}}^e(k)$. After this qualitative description and after explaining the main idea behind the calculations, we proceed with the exact calculations, some of which are moved to the Appendix.

Proof of Theorem 6: Consider the Chernoff bound on $\alpha_{i,\text{dis}}(k)$ given by eqn. (52), and denote by $\bar{m}_0 := \max_{i=1, \dots, N} |[m_\eta^{(0)}]_i|$. We first express the Chernoff upper bound on $\limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k)$ in terms of the “modes of decay,” similarly as in the Proof of Theorem 5. Namely, it can be shown (details

are in the Appendix) that the following inequality holds:

$$\begin{aligned} \limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) &= \limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) \\ &\leq \limsup_{k \rightarrow \infty} \frac{1}{k} \log \bar{\mathcal{C}}(k\mu) \bar{\delta}(k\mu), \quad \forall \mu > 0, \forall \epsilon \in (0, 1), \forall B = 1, 2, \dots \end{aligned} \quad (61)$$

where

$$\bar{\mathcal{C}}(k\mu) = \exp \left((k-1) \left(\frac{1}{2N^2} \sigma_L^2 \mu^2 + \frac{1}{N} m_L^{(0)} \mu \right) \right) \quad (62)$$

$$\begin{aligned} \bar{\delta}(k\mu) &= \sum_{j=0}^{\mathcal{J}} \exp \left(|2\mu^2 - \mu| \bar{m} \left\{ 2(j+1)B + (k - (j+1)B)N\sqrt{N}\epsilon \right\} \right) \\ &\quad \exp \left(\frac{\mu^2}{2} \|S^\eta\| \left\{ (j+1)B + (k - (j+1)B)N\epsilon^2 \right\} \right) \left(\frac{N^4 r^{jB}}{\epsilon^2} \right). \end{aligned} \quad (63)$$

Notice that the summands in eqn. (63) are the “modes of decay” that correspond to the events \mathcal{A}_{j+1} , $j = 1, \dots, \mathcal{J} + 1$. Next, after bounding from above by “the slowest mode,” we get:

$$\bar{\delta}(k\mu) \leq (\mathcal{J} + 1) \max_{j=0, \dots, \mathcal{J}} \exp \left(|2\mu^2 - \mu| \bar{m} \left\{ 2(j+1)B + (k - (j+1)B)N\sqrt{N}\epsilon \right\} \right) \quad (64)$$

$$\exp \left(\frac{\mu^2}{2} \|S^\eta\| \left\{ (j+1)B + (k - (j+1)B)N\epsilon^2 \right\} \right) \left(\frac{N^4 r^{jB}}{\epsilon^2} \right). \quad (65)$$

Introduce the variable $\theta = \theta(j) = \frac{jB}{k}$. We now replace the maximum over the discrete set $j = 0, \dots, \mathcal{J}$, in eqn. (64) by the supremum over $j \in [0, \mathcal{J}]$, i.e., $\theta \in [0, 1]$; we get:

$$\delta(k\mu) \leq (\mathcal{J} + 1) \sup_{\theta \in [0, 1]} \exp \left(|2\mu^2 - \mu| \bar{m} \left\{ 2k\theta(1 - N\sqrt{N}\epsilon) + B(2 - N\sqrt{N}\epsilon) + kN\sqrt{N}\epsilon \right\} \right) \quad (66)$$

$$\exp \left(\frac{\mu^2}{2} \|S^\eta\| \left\{ k\theta(1 - N\epsilon^2) + B(1 - N\epsilon^2) + k\epsilon^2 \right\} \right) \left(\frac{N^4 r^{jB}}{\epsilon^2} \right).$$

Taking the limsup as $k \rightarrow \infty$, and then letting $\epsilon \rightarrow 0$, we obtain the following inequality:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{C}(k\mu) \leq \sup_{\theta \in [0, 1]} \phi(\theta; \mu), \quad \forall \mu > 0 \quad (67)$$

$$\phi(\theta; \mu) = \frac{1}{2N^2} \sigma_L^2 \mu^2 + \frac{1}{N} m_L^{(0)} \mu + 2|2\mu^2 - \mu| \bar{m} \theta + \frac{\mu^2}{2} \|S^\eta\| \theta + \theta \log r. \quad (68)$$

Eqn. (67) gives a family of upper bounds, indexed by $\mu > 0$, on the quantity $\limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k)$; we seek the most aggressive bound, i.e., we take the infimum over $\mu > 0$.

We first discuss the values of the parameter r , for which there exists a range $\mu \in [\frac{1}{2}, \bar{\mu}]$ (where $\bar{\mu}$ is given by eqn. (50),) where the supremum in eqn. (67) is attained at $\theta = 0$. It can be shown that this

range is nonempty if and only if:

$$|\log r| > \frac{1}{8} \|S^\eta\|. \quad (69)$$

In view of eqns. (67), (68), and (69), we have the following inequality:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{C}'(k\mu) \leq \min_{\mu \in [\frac{1}{2}, \bar{\mu}]} \left\{ \frac{1}{2N^2} \sigma_L^2 \mu^2 + \frac{1}{N} m_L^{(0)} \mu \right\}. \quad (70)$$

The global minimum (on $\mu \in \mathbb{R}$) of the function $\mu \mapsto \frac{1}{2N^2} \sigma_L^2 \mu^2 + \frac{1}{N} m_L^{(0)} \mu$ is attained at $\mu^* = N/2$; and it can be shown that it equals the negative of the Chernoff information $-\mathbf{C}_{\text{tot}} = -\frac{1}{8} (m_1 - m_0)^\top S^{-1} (m_1 - m_0)$. Thus, a sufficient condition for $\limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) \leq \limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{C}'(k\mu) \leq -\mathbf{C}_{\text{tot}}$ is that $\mu^* = \frac{N}{2} \in [\frac{1}{2}, \bar{\mu}]$. By straightforward algebra, it can be shown that the latter condition translates into the following condition:

$$|\log r| \geq \frac{1}{8} N^2 \left(1 + \left(1 - \frac{1}{N} \right) K \right) \|S^\eta\|. \quad (71)$$

The analysis of the upper bound on $\alpha_{i,\text{dis}}(k)$ remains true for the upper bound on $\beta_{i,\text{dis}}(k)$; hence, we conclude that, under condition (71), we have: $\limsup_{k \rightarrow \infty} \frac{1}{k} \log \beta_{i,\text{dis}}(k) \leq -\mathbf{C}_{\text{tot}}$. Thus, under the condition (71), the following holds:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) = \max \left(\limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k), \limsup_{k \rightarrow \infty} \frac{1}{k} \log \beta_{i,\text{dis}}(k) \right) \leq -\mathbf{C}_{\text{tot}}. \quad (72)$$

On the other hand, by the Chernoff Lemma (Lemma 3,) we also know that

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) \geq -\mathbf{C}_{\text{tot}}. \quad (73)$$

By eqns. (72) and (73), we conclude that $\lim_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) = -\mathbf{C}_{\text{tot}}$, for the values of r that satisfy the condition (71). Hence, condition (71) is a sufficient condition for asymptotic optimality.

Consider now the values of r such that:

$$\frac{1}{8} \|S^\eta\| < |\log r| < \frac{1}{8} N^2 \left(1 + \left(1 - \frac{1}{N} \right) K \right) \|S^\eta\|. \quad (74)$$

In this range, $\frac{1}{2} < \bar{\mu} < \mu^* = N/2$, and the minimum of the function $\mu \mapsto \frac{1}{2N^2} \sigma_L^2 \mu^2 + \frac{1}{N} m_L^{(0)} \mu$ on $\mu \in [\frac{1}{2}, \bar{\mu}]$ is attained at $\mu = \bar{\mu}$. Hence, for the values of r in the range (74), we have the following bound:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) = \max \left(\limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k), \limsup_{k \rightarrow \infty} \frac{1}{k} \log \beta_{i,\text{dis}}(k) \right) \quad (75)$$

$$\leq \frac{1}{2N^2} \sigma_L^2 \bar{\mu}^2 + \frac{1}{N} m_L^{(0)} \bar{\mu}. \quad (76)$$

Finally, consider the values of r such that the condition in (69) does not hold, i.e., $\frac{1}{8}\|S^\eta\| \geq |\log r|$. We omit further details, but it can be shown that, in this case, the supremum in eqn. (67) is attained at $\theta = 0$ only for the values of μ in a subset of $[0, \frac{1}{2}]$; and it can be shown that $\limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k)$ can be bounded as given by eqn. (50). ■

B. Necessary condition for asymptotic optimality

We proceed with necessary conditions for asymptotic optimality, for the case when the sensors' observations are spatially uncorrelated, i.e., the noise covariance $S = \text{Diag}(\sigma_1^2, \dots, \sigma_N^2)$. Denote by $E_i(k)$ the event that, at time k , sensor i is connected to at least one of the remaining sensors in the network; that is,

$$E_i(k) := \left\{ \max_{j=1, \dots, N, j \neq i} W_{ij}(k) > 0 \right\}. \quad (77)$$

Further, denote by $P_i(k) = P_i = \mathbb{P}(E_i(k))$. We have the following result.

Theorem 10 (Necessary condition for asymptotic optimality) Consider the distributed detection algorithm in eqns. (18) and (19) with spatially uncorrelated sensors' observations, and let Assumption 4 hold. Then, a necessary condition for the asymptotic optimality of distributed detection at sensor i is:

$$|\log(1 - P_i)| > \mathbf{C}_{\text{tot}} - \mathbf{C}_i. \quad (78)$$

Proof: Consider $\alpha_{i,\text{dis}}(k)$, but remark that the same conclusions will hold for $\beta_{i,\text{dis}}(k)$, as $\alpha_{i,\text{dis}}(k) = \beta_{i,\text{dis}}(k)$. (See the proof of Theorem 5.) Further, using inequalities (7), we have:

$$\alpha_{i,\text{dis}}(k) = \mathbb{P}_0(x_i(k) > 0) \quad (79)$$

$$\geq \mathbb{P}_0\left(x_i(k) > 0 \mid \bigcap_{s=1}^k E_i(s)\right) \mathbb{P}\left(\bigcap_{s=1}^k E_i(s)\right) \quad (80)$$

$$\geq \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2k}\mathbf{C}_i}{1 + 2k\mathbf{C}_i} e^{-k\mathbf{C}_i} (1 - P_i)^k. \quad (81)$$

Thus, we have that

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) \geq -(\mathbf{C}_i + |\log(1 - P_i)|). \quad (82)$$

On the other hand, we know that

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) \geq \liminf_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) \geq -\mathbf{C}_{\text{tot}}. \quad (83)$$

The claim of Theorem 10 now follows from eqns. (82) and (83). ■

VII. SIMULATIONS

In this section, we corroborate by simulation examples our analytical findings on the asymptotic behavior of distributed detection over random networks. Namely, we demonstrate the “phase change” behavior of distributed detection with respect to the speed of network information flow, as predicted by Theorem 6. Also, we demonstrate that a sensor with poor connectedness to the rest of the network cannot be an optimal detector, as predicted by Theorem 10; moreover, its performance approaches the performance of an isolated sensor, i.e., a sensor that works as an individual detector, as connectedness becomes worse and worse.

Simulation setup. We consider a supergraph G with $N = 40$ nodes and $M = 247$ edges. Nodes are uniformly distributed on a unit square and nodes within distance less than a radius r are connected by an edge. As averaging weights, we use the standard time-varying Metropolis weights $W_{ij}(k)$, defined for $\{i, j\} \in E, i \neq j$, by $W_{ij}(k) = 1/(1 + \max(d_i(k), d_j(k)))$, if the link $\{i, j\}$ is online at time k , and 0 otherwise. The quantity $d_i(k)$ represents the number of neighbors (i.e., the degree) of node i at time k . Also, $W_{ii}(k) = 1 - \sum_{j \in \Omega_i(k)} W_{ij}(k)$, for all i , and $W_{ij}(k) \equiv 0$, for $i \neq j, \{i, j\} \notin E$. The link failures are spatially and temporally independent. Each link $\{i, j\} \in E$ has the same probability of formation, i.e., the probability of being online at a time, $q_{ij} = q$. This network and weight model satisfy Assumption 4.

We assume equal prior probabilities, $\mathbb{P}(H_0) = \mathbb{P}(H_1) = 0.5$. We set the $N \times 1$ signal vector under H_1 (respectively, H_0) to be $m_1 = 1$ (respectively, $m_0 = 0$.) We generate randomly the covariance matrix S , as follows. We generate: a $N \times N$ matrix M_S , with the entries drawn independently from $U[0, 1]$ —the uniform distribution on $[0, 1]$; we set $R_S = M_S M_S^\top$; we decompose R_S via the eigenvalue decomposition: $R_S = Q_S \Lambda_S Q_S^\top$; we generate a $N \times 1$ vector u_S with the entries drawn independently from $U[0, 1]$; finally, we set $S = \alpha_S Q_S \text{Diag}(u_S) Q_S^\top$, where $\alpha_S > 0$ is a parameter. For the optimal centralized detector, we evaluate $P_{\text{cen}}^e(k)$ by formula (6). For the distributed detector, we evaluate $P_{i,\text{dis}}^e(k)$ by Monte Carlo simulations with 20,000 sample paths (20,000 for each hypothesis $H_l, l = 0, 1$) of the running consensus algorithm.

Exponential rate of decay of the error probability vs. the speed of information flow. First, we examine the asymptotic behavior of distributed detection when the speed of network information flow varies, i.e., when r varies. (We recall that $r := \lambda_2(\mathbb{E}[W(k)^2])$.) To this end, we fix the supergraph, and then we vary the formation probability of links q from 0 to 0.75. Figure 3 (bottom left) plots the estimated exponential rate of decay, averaged across sensors, versus q . Figure 3 (bottom right) plots the same estimated exponential rate of decay versus $1 - r$. We can see that there is a “phase change”

behavior, as predicted by Theorem 6. For q greater than 0.1, i.e., for $1 - r > 0.25$, the rate of decay of error probability is approximately the same as for the optimal centralized detector \mathbf{C}_{tot} —the simulation estimate of \mathbf{C}_{tot} is 0.0106.³ For $q < 0.1$, i.e., for $1 - r < 0.25$, detection performance becomes worse and worse as q decreases. Figure 3 (top) plots the estimated error probability, averaged across sensors, for different values of q . We can see that the curves are “stretched” for small values of q ; after q exceeds a threshold (on the order of 0.1,) the curves cluster, and they have approximately the same slope (the error probability has approximately the same decay rate,) equal to the optimal slope.

Study of a sensor with poor connectivity to the rest of the network. Next, we demonstrate that a sensor with poor connectivity to the rest of the network cannot be an asymptotically optimal detector, as predicted by Theorem 10; its performance approaches the performance of an individual detector-sensor, when its connectivity becomes worse and worse. For i -th individual detector-sensor (no cooperation between sensors,) it is easy to show that the Bayes probability of error, $P_{i,\text{no cooper.}}^e(k)$ equals: $P_{i,\text{no cooper.}}^e(k) = \mathcal{Q}\left(\sqrt{k} \frac{m_{i,\text{no cooper.}}}{\sigma_{i,\text{no cooper.}}}\right)$, where $m_{i,\text{no cooper.}} = \frac{1}{2} \frac{[m_1]_i^2}{S_{ii}}$, and $\sigma_{i,\text{no cooper.}}^2 = \frac{[m_1]_i^2}{S_{ii}}$. It is easy to show that the Chernoff information (equal to $\lim_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{no cooper.}}^e(k)$) for sensor i , in the absence of cooperation, is given by $\frac{1}{8} \frac{[m_1]_i^2}{S_{ii}}$.

We now detail the simulation setup. We consider a supergraph with $N = 35$ nodes and $M = 263$ edges. We initially generate the supergraph as a geometric disc graph, but then we isolate sensor 35 from the rest of the network, by keeping it connected only to sensor 3. We then vary the formation probability of the link $\{3, 35\}$, $q_{3,35}$, from 0.05 to 0.5 (see Figure 4.) All other links in the supergraph have the formation probability of 0.8. Figure 4 plots the error probability for: 1) the optimal centralized detection; 2) distributed detection at each sensor, with cooperation (running consensus;) and 3) detection at each sensor, without cooperation (sensors do not communicate.) Figure 4 shows that, when $q_{3,35} = 0.05$, sensor 35 behaves almost as bad as the individual sensors, that do not communicate (cooperate) with each other. As q increases, the performance of sensor 35 gradually improves.

VIII. CONCLUSION

We studied distributed detection over random networks when at each time step k each sensor: 1) averages its decision variable with the neighbors decision variables; and 2) accounts on-the-fly for its new observation. We analyzed how asymptotic detection performance, i.e., the exponential decay rate of the error probability, depends on the random network connectivity, i.e., on the speed of information flow across

³In this numerical example, the theoretical value of \mathbf{C}_{tot} is 0.009. The estimated value shows an error because the decay of the error probability, for the centralized detection, and for distributed detection with a large q , tends to slow down slightly when k is very large; this effect is not completely captured by simulation with $k < 700$.

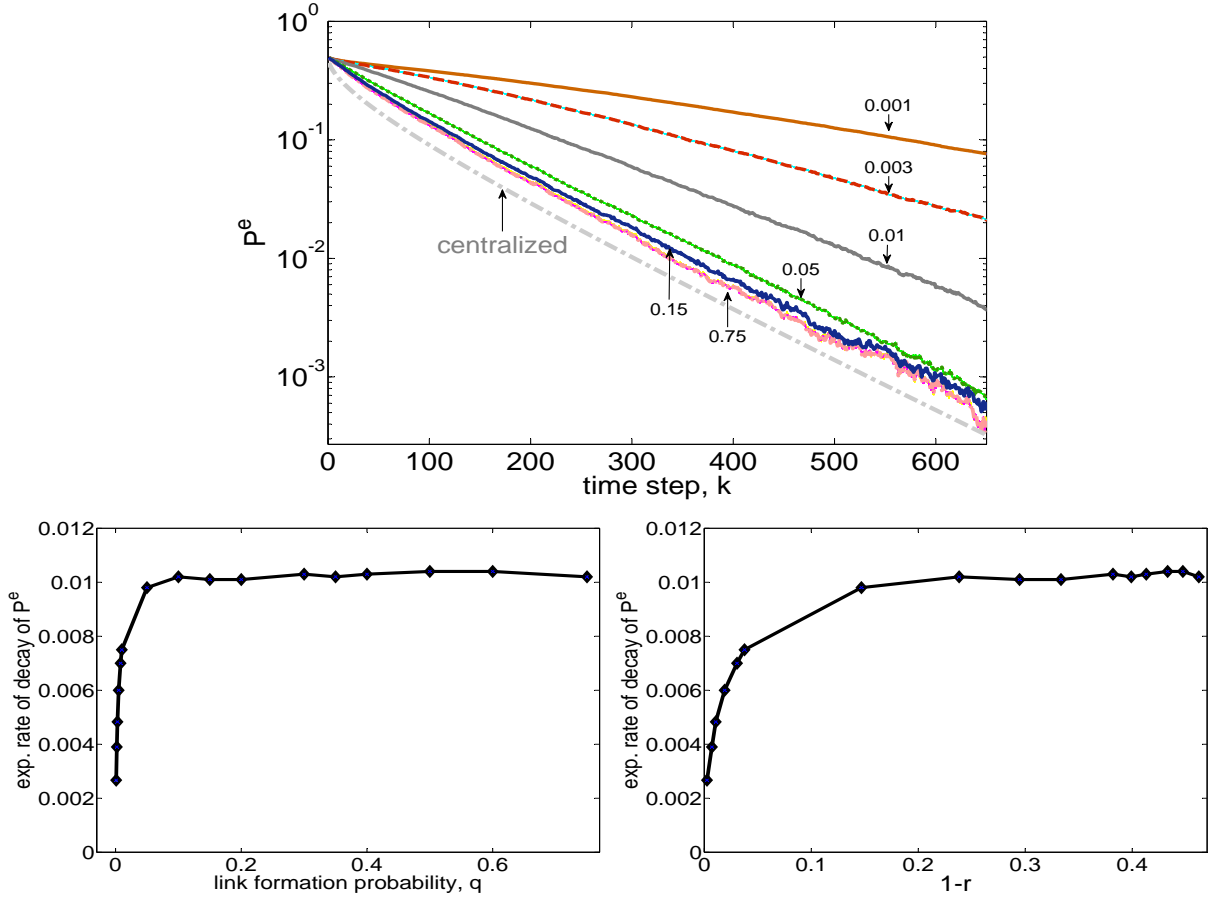


Fig. 3. Monte Carlo estimate of the performance of distributed detection for different values of the link formation probability q . Top: Error probability averaged across N sensors. Each line is labeled with the value of q ; performance of centralized detection is plotted in gray. Bottom left (respectively, right): Estimated exponential rate of decay of the error probability vs. q (respectively, $1 - r$). Recall that $r := \lambda_2 (\mathbb{E} [W(k)^2])$.

network. We showed that distributed detection exhibits a “phase change.” Namely, distributed detection is asymptotically optimal, if the network speed of information flow is above a Chernoff information dependent threshold. When below the threshold, we find a lower bound on the achievable performance (the exponential rate of decay of error probability,) as a function of the network connectivity and the Chernoff information. Simulation examples demonstrate our theoretical findings on the asymptotic performance of distributed detection.

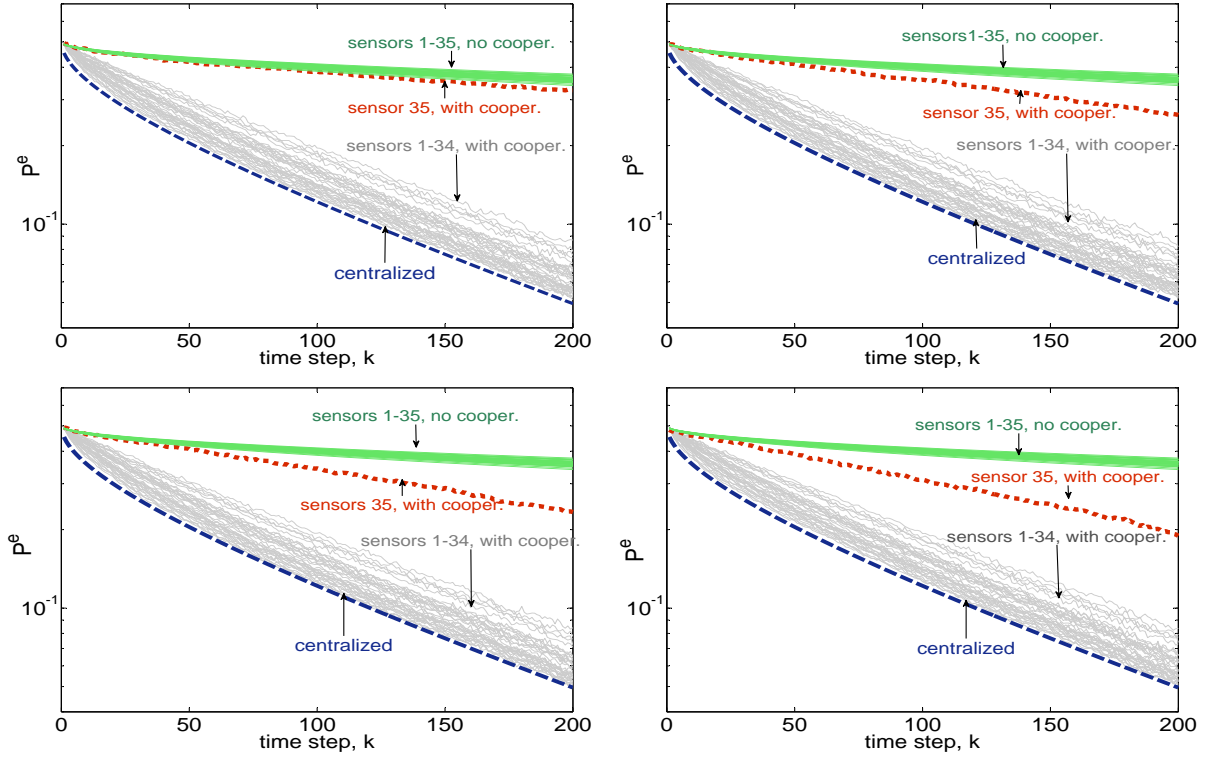


Fig. 4. Error probability averaged across sensors for the optimal centralized detection, distributed detection at each sensor (with cooperation), and detection at each sensor, without cooperation. The formation probability $q_{3,35}$ of the link $\{3, 35\}$ varies between 0.05 and 0.5: $q_{3,35}=0.05$ (top right); 0.2 (top left); 0.3 (bottom left); 0.5 (bottom right).

APPENDIX

A. Proof of inequalities (43)

Recall $\chi(l; k)$ in eqn. (38). Using eqn. (39), and noting that $(1 - p)^{k-1} \geq p(1 - p)^{k-1}$, we bound $\alpha_{i,\text{dis}}(k)$ from below as follows:

$$\alpha_{i,\text{dis}}(k) \geq \max_{l=0,\dots,k-1} \mathcal{Q}(\chi(l; k)) p(1 - p)^{k-l-1} \quad (84)$$

$$\geq \frac{1}{\sqrt{2\pi}} \max_{l=0,\dots,k-1} \frac{\chi(l; k)}{1 + \chi(l; k)^2} e^{-\frac{1}{2}\chi(l; k)^2} p(1 - p)^{k-l-1} \quad (85)$$

$$\geq \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2k\mathbf{C}_i}}{1 + 2kN\mathbf{C}_i} \max_{l=0,\dots,k-1} e^{-\frac{1}{2}\chi(l; k)^2} p(1 - p)^{k-l-1} \quad (86)$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2k\mathbf{C}_i}}{1 + 2kN\mathbf{C}_i} \max_{l=0,\dots,k-1} p e^{-\frac{k\mathbf{C}_{\text{tot}}}{1+(N-1)(1-\frac{1}{k})} - (k-l-1)|\log(1-p)|} \quad (87)$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2k\mathbf{C}_i}}{1 + 2kN\mathbf{C}_i} \max_{j=0,\dots,k-1} p e^{-\frac{k\mathbf{C}_{\text{tot}}}{1+(N-1)\frac{(j+1)}{k}} - j|\log(1-p)|} \quad (88)$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{2k\mathbf{C}_i}}{1 + 2kN\mathbf{C}_i} \max_{j=0,\dots,k-1} p e^{-k\phi(j; k)}. \quad (89)$$

Inequality in eqn. (85) is due to (7), equality in (88) is by letting $j = k - l - 1$, and equality in (89) is by definition of $\phi(j; k)$ in eqn. (41). Taking $\liminf_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k)$, and using (33), (84)-(89) yields:

$$\liminf_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) \geq \liminf_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) \geq - \limsup_{k \rightarrow \infty} \left\{ \min_{j=0, \dots, k-1} \phi(j; k) \right\}. \quad (90)$$

Next, using eqn. (39), we bound $\mathbb{P}_0(x_i(k) > 0)$ from above as follows:

$$\alpha_{i,\text{dis}}(k) \leq k \max_{l=0, \dots, k-1} \mathcal{Q}(\chi(l; k)) (1-p)^{k-l-1} \quad (91)$$

$$\leq k \frac{1}{\sqrt{2\pi}} \max_{l=0, \dots, k-1} \frac{1}{\chi(l; k)} e^{-\frac{1}{2}\chi(l; k)^2} (1-p)^{k-l-1} \quad (92)$$

$$\leq k \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\mathbf{C}_i k}} \max_{l=0, \dots, k-1} e^{-\frac{1}{2}\chi(l; k)^2} (1-p)^{k-l-1} \quad (93)$$

$$= k \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\mathbf{C}_i k}} \max_{l=0, \dots, k-1} e^{-\frac{k\mathbf{C}_{\text{tot}}}{1+(N-1)(1-\frac{1}{k})} - (k-l-1)|\log(1-p)|} \quad (94)$$

$$= k \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\mathbf{C}_i k}} \max_{j=0, \dots, k-1} e^{-\frac{k\mathbf{C}_{\text{tot}}}{1+(N-1)\frac{1}{k}} - j|\log(1-p)|} \quad (95)$$

$$= k \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\mathbf{C}_i k}} \max_{j=0, \dots, k-1} e^{-k\phi(j; k)}. \quad (96)$$

Inequality in (92) is due (7), and equality in (96) is by definition of $\phi(j; k)$ in eqn. (41). We complete the proof by noting that equality in (32) and inequalities (91)-(95) imply that:

$$\limsup_{k \rightarrow \infty} \frac{1}{k} \log P_{i,\text{dis}}^e(k) = \limsup_{k \rightarrow \infty} \frac{1}{k} \log \alpha_{i,\text{dis}}(k) \leq - \liminf_{k \rightarrow \infty} \left\{ \min_{j \in \{0, \dots, k-1\}} \phi(j; k) \right\}. \quad (97)$$

B. Proof of Lemma 8

Consider the random vector $z^{(i)}(k, j) = \tilde{\Phi}(k, j)e_i$, which is equal to the i -th column of the matrix $\tilde{\Phi}(k, j)$. First, by the Markov inequality, we have:

$$\mathbb{P} \left(\|z^{(i)}(k, j)\| > \epsilon \right) \leq \frac{1}{\epsilon^2} \mathbb{E} \left[z^{(i)}(k, j)^\top z^{(i)}(k, j) \right]. \quad (98)$$

Now, we can bound $\mathbb{E} [z^{(i)}(k, j)^\top z^{(i)}(k, j)]$ as follows:

$$\begin{aligned} \mathbb{E} [z^{(i)}(k, j)^\top z^{(i)}(k, j)] &= \mathbb{E} \left[\mathbb{E} \left[e_i^\top \tilde{\Phi}(k-1, j)^\top \tilde{W}(k-1) \tilde{W}(k-1) \tilde{\Phi}(k-1, j) e_i \mid \tilde{W}(j), \dots, \tilde{W}(k-2) \right] \right] \\ &= \mathbb{E} \left[e_i^\top \tilde{\Phi}(k-1, j)^\top \mathbb{E} [\tilde{W}(k-1) \tilde{W}(k-1)] \tilde{\Phi}(k-1, j) e_i \mid \tilde{W}(j), \dots, \tilde{W}(k-2) \right] \\ &\leq \|\mathbb{E} [\tilde{W}(k-1)^2]\| \cdot \mathbb{E} [e_i^\top \tilde{\Phi}(k-1, j)^\top \tilde{\Phi}(k-1, j) e_i] \\ &= r \mathbb{E} [z^{(i)}(k-1, j)^\top z^{(i)}(k-1, j)]. \end{aligned} \quad (99)$$

Now, bounding successively $\mathbb{E} [z^{(i)}(k-s, j)^\top z^{(i)}(k-s, j)]$ for $s = 1, \dots, k-j$, as in eqn. (99), we obtain:

$$\mathbb{E} [z^{(i)}(k, j)^\top z^{(i)}(k, j)] \leq r^{k-j}, \quad (100)$$

which yields (by eqn. (98):)

$$\mathbb{P} \left(\|z^{(i)}(k, j)\| > \epsilon \right) \leq \frac{1}{\epsilon^2} r^{k-j}, \quad i = 1, \dots, N. \quad (101)$$

Now, observe that $\|z^{(i)}(k, j)\|_1 \leq N \|z^{(i)}(k, j)\|$, and hence,

$$\mathbb{P} \left(\|z^{(i)}(k, j)\|_1 > \epsilon \right) \leq \mathbb{P} \left(\|z^{(i)}(k, j)\| > \frac{\epsilon}{N} \right).$$

Further, in view of the matrix norm inequality $\|B\| \leq \sqrt{N} \|B\|_1$, we have:

$$\begin{aligned} \mathbb{P} \left(\|\tilde{\Phi}(k, j)\| > \epsilon \right) &\leq \mathbb{P} \left(\|\tilde{\Phi}(k, j)\|_1 > \frac{\epsilon}{\sqrt{N}} \right) = \mathbb{P} \left(\max_{i=1, \dots, N} \|z^{(i)}(k, j)\|_1 > \frac{\epsilon}{\sqrt{N}} \right) \\ &= \mathbb{P} \left(\cup_{i=1}^N \{ \|z^{(i)}(k, j)\|_1 > \frac{\epsilon}{\sqrt{N}} \} \right) \leq \sum_{i=1}^N \mathbb{P} \left(\|z^{(i)}(k, j)\|_1 > \frac{\epsilon}{\sqrt{N}} \right) \leq \frac{N^4}{\epsilon^2} r^{k-j}. \end{aligned}$$

C. Proof of Lemma 9

First, note that, if \mathcal{A}_j occurred, we have that:

$$\|\tilde{\Phi}(k, s)\| \leq \epsilon, \quad \forall s \leq k-jB,$$

since $\|\tilde{\Phi}(k, s)\| = \|[\tilde{\Phi}(k, k-jB)\tilde{\Phi}(k-jB-1, s)]\| \leq \|\tilde{\Phi}(k, k-jB)\| \|\tilde{\Phi}(k-jB-1, s)\|$, and $\|\tilde{\Phi}(k, k-jB)\| \leq \epsilon$ and $\|\tilde{\Phi}(k-jB-1, s)\| \leq 1$. Inequality (59) holds true, because:

$$\sum_{l=1}^N |[\tilde{\Phi}(k, s)]_{il}| \leq N \max_{i, l=1, \dots, N} |[\tilde{\Phi}(k, s)]_{il}|,$$

and $\max_{i, l=1, \dots, N} |[\tilde{\Phi}(k, s)]_{il}| \leq \sqrt{N} \|\tilde{\Phi}(k, s)\| \leq \sqrt{N}\epsilon$. Inequality (60) holds true, because, for $s \leq k-jB$,

$$\sum_{l=1}^N |[\tilde{\Phi}(k, s)]_{il}|^2 \leq \|\tilde{\Phi}(k, s)\|_F^2 \leq N \|\tilde{\Phi}(k, s)\|^2 = N\epsilon^2.$$

D. Proof of eqn. (61)

Consider the Chernoff bound on $\alpha_{i, \text{dis}}(k)$, given by:

$$\mathcal{C}(k\mu) := \mathbb{E}_0 [\exp(k\mu x_i(k))] = \mathbb{E}_0 \left[\exp \left(k\lambda^\top x(k) \right) \right], \quad (102)$$

where $\lambda = \mu e_i$, $\lambda \in \mathbb{R}^N$.

Further, we have:

$$\begin{aligned}\mathcal{C}(k\mu) &= \mathbb{E}_0 \left[\exp \left(\lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) + \lambda^\top \eta(k) \right) \right] \\ &= \mathbb{E}_0 \left[\exp \left(\lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) \right) \right] \mathbb{E}_0 \left[\exp \left(\lambda^\top \eta(k) \right) \right],\end{aligned}\tag{103}$$

where the last equality holds because $\eta(k)$ is independent from $\eta(j)$ and $W(j)$, $j = 1, \dots, k-1$. We will be interested in computing $\limsup_{k \rightarrow \infty} \frac{1}{k} \Lambda_k^{(l)}(k\mu)$, for all $\mu \in \mathbb{R}$; with this respect, remark that

$$\lim_{k \rightarrow \infty} \frac{1}{k} \log \mathbb{E}_0 \left[\exp \left(\lambda^\top \eta(k) \right) \right] = 0,$$

for all $\lambda \in \mathbb{R}^N$, because $\eta(k)$ is a Gaussian random variable and hence it has finite log-moment generating function at any point λ .

Thus, we have that $\limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{C}(k\mu) = \limsup_{k \rightarrow \infty} \frac{1}{k} \log \mathcal{C}'(k\mu)$, where

$$\mathcal{C}'(k\mu) = \mathbb{E}_0 \left[\exp \left(\lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) \right) \right].$$

We thus proceed with the computation of $\mathcal{C}'(k\mu)$. Conditioned on $W(1), W(2), \dots, W(k-1)$, the random variables $\lambda^\top \Phi(k, j) \eta(j)$, $j = 1, \dots, k-1$, are independent; moreover, they are Gaussian random variables, as linear transformation of the Gaussian variables $\eta(j)$. Recall that $m_\eta^{(0)}$ and S^η denote the mean and the covariance of $\eta(k)$ under hypothesis H_0 . After conditioning on $W(1), W(2), \dots, W(k-1)$, using the independence of $\eta(j)$ and $\eta(s)$ $s \neq j$, and using the expression for the moment generating function of $\eta(j)$, we obtain successively:

$$\begin{aligned}\mathcal{C}'(\mu) &= \mathbb{E} \left[\mathbb{E}_0 \left[\exp \left(\lambda^\top \sum_{j=1}^{k-1} \Phi(k, j) \eta(j) \right) \mid W(1), \dots, W(k-1) \right] \right] \\ &= \mathbb{E} \left[\mathbb{E}_0 \left[\prod_{j=1}^{k-1} \exp \left(\lambda^\top \Phi(k, j) \eta(j) \right) \mid W(1), \dots, W(k-1) \right] \right] \\ &= \mathbb{E} \left[\prod_{j=1}^{k-1} \mathbb{E} \left[\exp \left(\lambda^\top \Phi(k, j) m_\eta^{(l)} \right) \exp \left(\frac{1}{2} \lambda^\top \Phi(k, j)^\top S^\eta \Phi(k, j) \lambda \right) \right] \right] \\ &= \mathbb{E} \left[\prod_{j=1}^{k-1} \exp \left(\lambda^\top \left(\tilde{\Phi}(k, j) + J \right) m_\eta^{(0)} \right) \exp \left(\frac{1}{2} \left(J + \tilde{\Phi}(k, j) \right)^\top S^\eta \left(J + \tilde{\Phi}(k, j) \right) \right) \right].\end{aligned}$$

Denote further:

$$\delta(k\mu) := \mathbb{E} \left[\exp \left(\lambda^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) m_\eta^{(l)} \right) \exp \left(\frac{1}{2} \lambda^\top \left(\sum_{j=1}^{k-1} \tilde{\Phi}(k, j)^\top S^\eta \tilde{\Phi}(k, j) \right) \lambda \right) \right] \quad (104)$$

$$\exp \left(\frac{1}{2} \lambda^\top J S^\eta \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) \lambda \right) \exp \left(\frac{1}{2} \lambda^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j)^\top S^\eta J \lambda \right),$$

$$\bar{\mathcal{C}}(k\mu) := \exp \left((k-1) \left(\lambda^\top J m_\eta^{(l)} - \frac{1}{2} \lambda^\top J S^\eta J \lambda \right) \right), \quad (105)$$

where dependence on H_0 is dropped in the definition of $\delta(k\mu)$. Then, it is easy to see that $\mathcal{C}'(k\mu) = \bar{\mathcal{C}}(k\mu)\delta(k\mu)$.

Recall the expressions for v , $m_L^{(0)}$ and σ_L^2 in eqns. (10), (12). After straightforward algebra, it can be shown that $\bar{\mathcal{C}}(k\mu)$ equals the expression in eqn. (62). Remark further that $\delta(k\mu)$ equals:

$$\delta(k\mu) = \mathbb{E} \left[\exp \left((\mu - \mu^2) e_i^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) m_\eta^{(0)} \right) \exp \left(\frac{\mu^2}{2} e_i^\top \sum_{j=1}^{k-1} \tilde{\Phi}(k, j) S^\eta \tilde{\Phi}(k, j)^\top e_i \right) \right]. \quad (106)$$

Recall that $\bar{m}_0 := \max_{i=1, \dots, N} |[m_\eta^{(0)}]_i|$. Now, $\delta(k\mu)$ can be bounded from above as follows:

$$\begin{aligned} \delta(k\mu) &\leq \mathbb{E} \left[\exp \left(|\mu - \mu^2| \bar{m}_0 \sum_{j=1}^{k-1} \sum_{l=1}^N |[\tilde{\Phi}(k, j)]_{il}| \right) \exp \left(\frac{\mu^2}{2} \|S^\eta\| \sum_{j=1}^{k-1} |[\tilde{\Phi}(k, j)]_{il}|^2 \right) \right] \\ &\leq \mathbb{E} \left[\exp \left(|\mu - \mu^2| \bar{m}_0 \sum_{j=1}^k \sum_{l=1}^N |[\tilde{\Phi}(k, j)]_{il}| \right) \exp \left(\frac{\mu^2}{2} \|S^\eta\| \sum_{j=1}^k |[\tilde{\Phi}(k, j)]_{il}|^2 \right) \right]. \end{aligned} \quad (107)$$

Using Lemmas 9, 8 and 7, using the total probability with the partition \mathcal{A}_j , $j = 1, \dots, \mathcal{J} + 1$, we get the inequality (61).

REFERENCES

- [1] R. Viswanatan and P. R. Varshney, "Decentralized detection with multiple sensors: Part I—fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, January 1997.
- [2] J. F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.
- [3] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Processing*, vol. 2, pp. 297–344, 1993.
- [4] R. S. Blum, S. A. Kassam, and H. V. Poor, "Decentralized detection with multiple sensors: Part II—advanced topics," *Proc. IEEE*, vol. 85, pp. 64–79, January 1997.
- [5] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, June 2008.
- [6] S. Kar and J. M. F. Moura, "Consensus based detection in sensor networks: topology design under practical constraints," in *Proc. Workshop Inf. Theory in Sensor Networks*, Santa Fe, NM, June 2007.

- [7] P. Braca, S. Marano, V. Matta, and P. Willet, "Asymptotic optimality of running consensus in testing binary hypothesis," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814–825, February 2010.
- [8] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS-based detection over adaptive networks," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, October 2009.
- [9] —, "Distributed detection over adaptive networks based on diffusion estimation schemes," in *Proc. IEEE SPAWC '09, 10th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Perugia, Italy, June 2009, pp. 61–65.
- [10] D. Jakovetic, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3788–3801, July 2010.
- [11] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.
- [12] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [13] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, March 2010.
- [14] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. on Signal Processing*, vol. 57, no. 6, pp. 2365–2381, June 2009.
- [15] —, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Trans. on Signal Processing*, vol. 57, no. 11, pp. 4583–4588, November 2009.
- [16] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," August 2008, submitted for publication, 51 pages. [Online]. Available: arXiv:0809.0009v1 [cs.MA]
- [17] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. New York: Springer-Verlag, 1987.
- [18] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection over time varying networks: large deviations analysis," in *48th Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2010.
- [19] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley and Sons, 1991.
- [20] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: the saddlepoint approximation," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 327–340, January 2007.
- [21] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Boston, MA: Jones and Barlett, 1993.
- [22] H. L. V. Trees, *Detection, Estimation, and Modulation Theory - Part I - Detection, Estimation, and Linear Modulation Theory*. John Wiley and Sons, 2001.
- [23] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3375–3380, July 2008.